NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL LIBRARIES

This application is a continuation-in-part of United States provisional application 06/198,069, filed April 17, 2000, a continuation-in-part of United States patent application 09/837,306, filed on April 17, 2001, and a continuation-in-part of United States application XX/XXX,XXX, filed by Express Mail(EI125454535US) on October 25, 2001. All of the earlier applications are specifically incorporated by reference herein.

The present invention relates to libraries of genetic packages that display and/or express a member of a diverse family of peptides, polypeptides or proteins and collectively display and/or express at least a portion of the diversity of the family. In an alternative embodiment, the invention relates to libraries that include a member of a diverse family of peptides, polypeptides or proteins and collectively comprise at least a portion of the diversity of the family. In a preferred embodiment, the displayed and/or expressed polypeptides are human Fabs.

More specifically, the invention is directed to the methods of cleaving single-stranded nucleic acids at chosen locations, the cleaved nucleic acids encoding, at least in part, the peptides, polypeptides

or proteins displayed on the genetic packages of, and/or expressed in, the libraries of the invention. In a preferred embodiment, the genetic packages are filamentous phage or phagemids or yeast.

The present invention further relates to vectors for displaying and/or expressing a diverse family of peptides, polypeptides or proteins.

The present invention further relates to methods of screening the libraries of the invention and to the peptides, polypeptides and proteins identified by such screening.

BACKGROUND OF THE INVENTION

It is now common practice in the art to prepare libraries of genetic packages that display, express or comprise a member of a diverse family of peptides, polypeptides or proteins and collectively display, express or comprise at least a portion of the diversity of the family. In many common libraries, the peptides, polypeptides or proteins are related to antibodies. Often, they are Fabs or single chain antibodies.

In general, the DNAs that encode members of the families to be displayed and/or expressed must be amplified before they are cloned and used to display and/or express the desired member. Such amplification typically makes use of forward and backward primers.

Such primers can be complementary to sequences native to the DNA to be amplified or complementary to oligonucleotides attached at the 5' or 3' ends of that DNA. Primers that are complementary to sequences native to the DNA to be amplified are disadvantaged in that they bias the members of the

families to be displayed. Only those members that contain a sequence in the native DNA that is substantially complementary to the primer will be amplified. Those that do not will be absent from the family. For those members that are amplified, any diversity within the primer region will be suppressed.

For example, in European patent 368,684 B1, the primer that is used is at the 5' end of the V_H region of an antibody gene. It anneals to a sequence region in the native DNA that is said to be "sufficiently well conserved" within a single species. Such primer will bias the members amplified to those having this "conserved" region. Any diversity within this region is extinguished.

It is generally accepted that human antibody genes arise through a process that involves a combinatorial selection of V and J or V, D, and J followed by somatic mutations. Although most diversity occurs in the Complementary Determining Regions (CDRs), diversity also occurs in the more conserved Framework Regions (FRs) and at least some of this diversity confers or enhances specific binding to antigens (Ag). As a consequence, libraries should contain as much of the CDR and FR diversity as possible.

To clone the amplified DNAs of the peptides, polypeptides or proteins that they encode for display on a genetic package and/or for expression, the DNAs must be cleaved to produce appropriate ends for ligation to a vector. Such cleavage is generally effected using restriction endonuclease recognition sites carried on the primers. When the primers are at the 5' end of DNA produced from reverse transcription of RNA, such restriction leaves deleterious 5' untranslated regions in the amplified DNA. These

regions interfere with expression of the cloned genes and thus the display of the peptides, polypeptides and proteins coded for by them.

SUMMARY OF THE INVENTION

It is an object of this invention to provide novel methods for constructing libraries that display, express or comprise a member of a diverse family of peptides, polypeptides or proteins and collectively display, express or comprise at least a portion of the diversity of the family. These methods are not biased toward DNAs that contain native sequences that are complementary to the primers used for amplification. They also enable any sequences that may be deleterious to expression to be removed from the amplified DNA before cloning and displaying and/or expressing.

It is another object of this invention to provide a method for cleaving single-stranded nucleic acid sequences at a desired location, the method comprising the steps of:

(i) contacting the nucleic acid with a single-stranded oligonucleotide, the oligonucleotide being functionally complementary to the nucleic acid in the region in which cleavage is desired and including a sequence that with its complement in the nucleic acid forms a restriction endonuclease recognition site that on restriction results in cleavage of the nucleic acid at the desired location; and

(ii) cleaving the nucleic acid solely at the recognition site formed by the complementation of the nucleic acid and the oligonucleotide;

the contacting and the cleaving steps being performed at a temperature sufficient to maintain the nucleic acid in substantially single-stranded form, the oligonucleotide being functionally complementary to the nucleic acid over a large enough region to allow the two strands to associate such that cleavage may occur at the chosen temperature and at the desired location, and the cleavage being carried out using a restriction endonuclease that is active at the chosen temperature.

It is a further object of this invention to provide an alternative method for cleaving single-stranded nucleic acid sequences at a desired location, the method comprising the steps of:

(i) contacting the nucleic acid with a partially double-stranded oligonucleotide, the single-stranded region of the oligonucleotide being functionally complementary to the nucleic acid in the region in which cleavage is desired, and the double-stranded region of the oligonucleotide having a restriction endonuclease recognition site; and

(ii) cleaving the nucleic acid solely at the cleavage site formed by the complementation of the nucleic acid and the single-stranded region of the oligonucleotide;

the contacting and the cleaving steps being performed at a temperature sufficient to maintain the nucleic acid in substantially single-stranded form, the oligonucleotide being functionally complementary to the nucleic acid over a large enough region to allow the two strands to associate such that cleavage may occur

at the chosen temperature and at the desired location, and the cleavage being carried out using a restriction endonuclease that is active at the chosen temperature.

In an alternative embodiment of this object of the invention, the restriction endonuclease recognition site is not initially located in the double-stranded part of the oligonucleotide. Instead, it is part of an amplification primer, which primer is complementary to the double-stranded region of the oligonucleotide. On amplification of the DNA-partially double-stranded combination, the restriction endonuclease recognition site carried on the primer becomes part of the DNA. It can then be used to cleave the DNA.

Preferably, the restriction endonuclease recognition site is that of a Type II-S restriction endonuclease whose cleavage site is located at a known distance from its recognition site.

It is another object of the present invention to provide a method of capturing DNA molecules that comprise a member of a diverse family of DNAs and collectively comprise at least a portion of the diversity of the family. These DNA molecules in single-stranded form have been cleaved by one of the methods of this invention. This method involves ligating the individual single-stranded DNA members of the family to a partially duplex DNA complex. The method comprises the steps of:

(i) contacting a single-stranded nucleic acid sequence that has been cleaved with a restriction endonuclease with a partially double-stranded oligonucleotide, the single-stranded region of the oligonucleotide being functionally complementary to the nucleic

acid in the region that remains after cleavage, the double-stranded region of the oligonucleotide including any sequences necessary to return the sequences that remain after cleavage into proper reading frame for expression and containing a restriction endonuclease recognition site 5' of those sequences; and

(ii) cleaving the partially double-stranded oligonucleotide sequence solely at the restriction endonuclease cleavage site contained within the double-stranded region of the partially double-stranded oligonucleotide.

As before, in this object of the invention, the restriction endonuclease recognition site need not be located in the double-stranded portion of the oligonucleotide. Instead, it can be introduced on amplification with an amplification primer that is used to amplify the DNA-partially double-stranded oligonucleotide combination.

It is another object of this invention to prepare libraries, that display, express or comprise a diverse family of peptides, polypeptides or proteins and collectively display, express or comprise at least part of the diversity of the family, using the methods and DNAs described above.

It is an object of this invention to screen those libraries to identify useful peptides, polypeptides and proteins and to use those substances in human therapy.

Additional objects of the invention are reflected in claims 1-116. Each of these claims is

specifically incorporated by reference in this specification.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic of various methods that may be employed to amplify VH genes without using primers specific for VH sequences.

FIG. 2 is a schematic of various methods that may be employed to amplify VL genes without using primers specific for VL sequences.

FIG. 3 is a schematic of RACE amplification of antibody heavy and light chains.

FIG. 4 depicts gel analysis of amplification products obtained after the primary PCR reaction from 4 different patient samples.

FIG. 5 depicts gel analysis of cleaved kappa DNA from Example 2.

FIG. 6 depicts gel analysis of extender-cleaved kappa DNA from Example 2.

FIG. 7 depicts gel analysis of the PCR product from the extender-kappa amplification from Example 2.

FIG. 8 depicts gel analysis of purified PCR product from the extender-kappa amplification from Example 2.

FIG. 9 depicts gel analysis of cleaved and ligated kappa light chains from Example 2.

FIG. 10 is a schematic of the design for CDR1 and CDR2 synthetic diversity.

FIG. 11 is a schemaitc of the cloning schedule for construction of the heavy chain repertoire.

FIG. 12 is a schematic of the cleavage and ligation of the antibody light chain.

FIG. 13 depicts gel analysis of cleaved and ligated lambda light chains from Example 4.

FIG. 14 is a schematic of the cleavage and ligation of the antibody heavy chain.

FIG. 15 depicts gel analysis of cleaved and ligated lambda light chains from Example 5.

FIG. 16 is a schematic of a phage display vector.

FIG. 17 is a schematic of a Fab cassette.

FIG. 18 is a schematic of a process for incorporating fixed FR1 residues in an antibody lambda sequence.

FIG. 19 is a schematic of a process for incorporating fixed FR1 residues in an antibody kappa sequence.

FIG. 20 is a schematic of a process for incorporating fixed FR1 residues in an antibody heavy chain sequence.

TERMS

In this application, the following terms and abbreviations are used:

Sense strand	The upper strand of ds DNA as usually written. In the sense strand, 5'-ATG-3' codes for Met.
Antisense strand	The lower strand of ds DNA as usually written. In the antisense strand, 3'-TAC-5' would correspond to a Met codon in the sense strand.

Forward primer	A "forward" primer is complementary to a part of the sense strand and primes for synthesis of a new antisense-strand molecule. "Forward primer" and "lower-strand primer" are equivalent.
Backward primer	A "backward" primer is complementary to a part of the antisense strand and primes for synthesis of a new sense-strand molecule. "Backward primer" and "top-strand primer" are equivalent.
Bases	Bases are specified either by their position in a vector or gene as their position within a gene by codon and base. For example, "89.1" is the first base of codon 89, 89.2 is the second base of codon 89.
Sv	Streptavidin
Ap	Ampicillin
ap^R	A gene conferring ampicillin resistance.
RERS	Restriction endonuclease recognition site

RE	Restriction endonuclease - cleaves preferentially at RERS
URE	Universal restriction endonuclease
Functionally complementary	Two sequences are sufficiently complementary so as to anneal under the chosen conditions.
AA	Amino acid
PCR	Polymerization chain reaction
GLGs	Germline genes
Ab	Antibody: an immunoglobin. The term also covers any protein having a binding domain which is homologous to an immunoglobin binding domain. A few examples of antibodies within this definition are, *inter alia*, immunoglobin isotypes and the Fab, $F(ab^1)_2$, scfv, Fv, dAb and Fd fragments.
Fab	Two chain molecule comprising an Ab light chain and part of a heavy-chain.

scFv	A single-chain Ab comprising either VH::linker::VL or VL::linker::VH
w.t.	Wild type
HC	Heavy chain
LC	Light chain
VK	A variable domain of a Kappa light chain.
VH	A variable domain of a heavy chain.
VL	A variable domain of a lambda light chain.

In this application when it is said that nucleic acids are cleaved solely at the cleavage site of a restriction endonuclease, it should be understood that minor cleavage may occur at random, *e.g.*, at non-specific sites other than the specific cleavage site that is characteristic of the restriction endonuclease. The skilled worker will recognize that such non-specific, random cleavage is the usual occurrence. Accordingly, "solely at the cleavage site" of a restriction endonuclease means that cleavage occurs preferentially at the site characteristic of that endonuclease.

As used in this application and claims, the term "cleavage site formed by the complementation of the nucleic acid and the single-stranded region of the

oligonucleotide" includes cleavage sites formed by the single-stranded portion of the partially double-stranded ologonucleotide duplexing with the single-stranded DNA, cleavage sites in the double-stranded portion of the partially double-stranded oligonucleotide, and cleavage sites introduced by the amplification primer used to amplify the single-stranded DNA-partially double-stranded oligonucleotide combination.

In the two methods of this invention for preparing single-stranded nucleic acid sequences, the first of those cleavage sites is preferred. In the methods of this invention for capturing diversity and cloning a family of diverse nucleic acid sequences, the latter two cleavage sites are preferred.

In this application, all references referred to are specifically incorporated by reference.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The nucleic acid sequences that are useful in the methods of this invention, *i.e.*, those that encode at least in part the individual peptides, polypeptides and proteins displayed, or expressed in or comprising the libraries of this invention, may be native, synthetic or a combination thereof. They may be mRNA, DNA or cDNA. In the preferred embodiment, the nucleic acids encode antibodies. Most preferably, they encode Fabs.

The nucleic acids useful in this invention may be naturally diverse, synthetic diversity may be introduced into those naturally diverse members, or the diversity may be entirely synthetic. For example, synthetic diversity can be introduced into one or more CDRs of antibody genes. Preferably, it is introduced

into CDR1 and CDR2 of immunoglobulins. Preferably, natural diversity is captured in the CDR3 regions of the immunoglogin genes of this invention from B cells. Most preferably, the nucleic acids of this invention comprise a population of immunoglobin genes that comprise synthetic diversity in at least one, and more preferably both of the CDR1 and CDR2 and diversity in CDR3 captured from B cells.

Synthetic diversity may be created, for example, through the use of TRIM technology (U.S. 5,869,644). TRIM technology allows control over exactly which amino-acid types are allowed at variegated positions and in what proportions. In TRIM technology, codons to be diversified are synthesized using mixtures of trinucleotides. This allows any set of amino acid types to be included in any proportion.

Another alternative that may be used to generate diversified DNA is mixed oligonucleotide synthesis. With TRIM technology, one could allow Ala and Trp. With mixed oligonucleotide synthesis, a mixture that included Ala and Trp would also necessarily include Ser and Gly. The amino-acid types allowed at the variegated positions are picked with reference to the structure of antibodies, or other peptides, polypeptides or proteins of the family, the observed diversity in germline genes, the observed somatic mutations frequently observed, and the desired areas and types of variegation.

In a preferred embodiment of this invention, the nucleic acid sequences for at least one CDR or other region of the peptides, polypeptides or proteins of the family are cDNAs produced by reverse transcription from mRNA. More preferably, the mRNAs are obtained from peripheral blood cells, bone marrow

cells, spleen cells or lymph node cells (such as B-lymphocytes or plasma cells) that express members of naturally diverse sets of related genes. More preferable, the mRNAs encode a diverse family of antibodies. Most preferably, the mRNAs are obtained from patients suffering from at least one autoimmune disorder or cancer. Preferably, mRNAs containing a high diversity of autoimmune diseases, such as systemic lupus erythematosus, systemic sclerosis, rheumatoid arthritis, antiphospholipid syndrome and vasculitis are used.

In a preferred embodiment of this invention, the cDNAs are produced from the mRNAs using reverse transcription. In this preferred embodiment, the mRNAs are separated from the cell and degraded using standard methods, such that only the full length (*i.e.*, capped) mRNAs remain. The cap is then removed and reverse transcription used to produce the cDNAs.

The reverse transcription of the first (antisense) strand can be done in any manner with any suitable primer. See, *e.g.*, HJ de Haard et al., Journal of Biological Chemistry, 274(26):18218-30 (1999). In the preferred embodiment of this invention where the mRNAs encode antibodies, primers that are complementary to the constant regions of antibody genes may be used. Those primers are useful because they do not generate bias toward subclasses of antibodies. In another embodiment, poly-dT primers may be used (and may be preferred for the heavy-chain genes). Alternatively, sequences complementary to the primer may be attached to the termini of the antisense strand.

In one preferred embodiment of this invention, the reverse transcriptase primer may be biotinylated, thus allowing the cDNA product to be

immobilized on streptavidin (Sv) beads. Immobilization can also be effected using a primer labeled at the 5' end with one of a) free amine group, b) thiol, c) carboxylic acid, or d) another group not found in DNA that can react to form a strong bond to a known partner on an insoluble medium. If, for example, a free amine (preferably primary amine) is provided at the 5' end of a DNA primer, this amine can be reacted with carboxylic acid groups on a polymer bead using standard amide-forming chemistry. If such preferred immobilization is used during reverse transcription, the top strand RNA is degraded using well-known enzymes, such as a combination of RNAseH and RNAseA, either before or after immobilization.

The nucleic acid sequences useful in the methods of this invention are generally amplified before being used to display and/or express the peptides, polypeptides or proteins that they encode. Prior to amplification, the single-stranded DNAs may be cleaved using either of the methods described before. Alternatively, the single-stranded DNAs may be amplified and then cleaved using one of those methods.

Any of the well known methods for amplifying nucleic acid sequences may be used for such amplification. Methods that maximize, and do not bias, diversity are preferred. In a preferred embodiment of this invention where the nucleic acid sequences are derived from antibody genes, the present invention preferably utilizes primers in the constant regions of the heavy and light chain genes and primers to a synthetic sequence that are attached at the 5' end of the sense strand. Priming at such synthetic sequence avoids the use of sequences within the variable regions of the antibody genes. Those variable region priming

sites generate bias against V genes that are either of rare subclasses or that have been mutated at the priming sites. This bias is partly due to suppression of diversity within the primer region and partly due to lack of priming when many mutations are present in the region complementary to the primer. The methods disclosed in this invention have the advantage of not biasing the population of amplified antibody genes for particular V gene types.

The synthetic sequences may be attached to the 5' end of the DNA strand by various methods well known for ligating DNA sequences together. RT CapExtention is one preferred method.

In RT CapExtention (derived from Smart PCR(TM)), a short overlap (5'-...GGG-3' in the upper-strand primer (USP-GGG) complements 3'-CCC....5' in the lower strand) and reverse transcriptases are used so that the reverse complement of the upper-strand primer is attached to the lower strand.

FIGs. 1 and 2 show schematics to amplify VH and VL genes using RT CapExtention. FIG. 1 shows a schematic of the amplification of VH genes. FIG. 1, Panel A shows a primer specific to the poly-dT region of the 3' UTR priming synthesis of the first, lower strand. Primers that bind in the constant region are also suitable. Panel B shows the lower strand extended at its 3' end by three Cs that are not complementary to the mRNA. Panel C shows the result of annealing a synthetic top-strand primer ending in three GGGs that hybridize to the 3' terminal CCCs and extending the reverse transcription extending the lower strand by the reverse complement of the synthetic primer sequence. Panel D shows the result of PCR amplification using a 5' biotinylated synthetic top-strand primer that

replicates the 5' end of the synthetic primer of panel C and a bottom-strand primer complementary to part of the constant domain. Panel E shows immobilized double-stranded (ds) cDNA obtained by using a 5'-biotinylated top-strand primer.

FIG. 2 shows a similar schematic for amplification of VL genes. FIG. 2, Panel A shows a primer specific to the constant region at or near the 3' end priming synthesis of the first, lower strand. Primers that bind in the poly-dT region are also suitable. Panel B shows the lower strand extended at its 3' end by three Cs that are not complementary to the mRNA. Panel C shows the result of annealing a synthetic top-strand primer ending in three GGGs that hybridize to the 3' terminal CCCs and extending the reverse transcription extending the lower strand by the reverse complement of the synthetic primer sequence. Panel D shows the result of PCR amplification using a 5' biotinylated synthetic top-strand primer that replicates the 5' end of the synthetic primer of panel C and a bottom-strand primer complementary to part of the constant domain. The bottom-strand primer also contains a useful restriction endonuclease site, such as *Asc*I. Panel E shows immobilized ds cDNA obtained by using a 5'-biotinylated top-strand primer.

In FIGs. 1 and 2, each V gene consists of a 5' untranslated region (UTR) and a secretion signal, followed by the variable region, followed by a constant region, followed by a 3' untranslated region (which typically ends in poly-A). An initial primer for reverse transcription may be complementary to the constant region or to the poly A segment of the 3'-UTR. For human heavy-chain genes, a primer of 15 T is preferred. Reverse transcriptases attach several C

residues to the 3' end of the newly synthesized DNA. RT CapExtention exploits this feature. The reverse transcription reaction is first run with only a lower-strand primer. After about 1 hour, a primer ending in GGG (USP-GGG) and more RTase are added. This causes the lower-strand cDNA to be extended by the reverse complement of the USP-GGG up to the final GGG. Using one primer identical to part of the attached synthetic sequence and a second primer complementary to a region of known sequence at the 3' end of the sense strand, all the V genes are amplified irrespective of their V gene subclass.

In another preferred embodiment, synthetic sequences may be added by Rapid Amplification of cDNA Ends (RACE) (see Frohman, M.A., Dush, M.K., & Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA (85): 8998-9002).

FIG. 1 shows a schematic of RACE amplification of antibody heavy and light chains. First, mRNA is selected by treating total or poly(A+) RNA with calf intestinal phosphatase (CIP) to remove the 5'-phosphate from all molecules that have them such as ribosomal RNA, fragmented mRNA, tRNA and genomic DNA. Full length mRNA (containing a protective 7-methyl cap structure) is uneffected. The RNA is then treated with tobacco acid pyrophosphatase (TAP) to remove the cap structure from full length mRNAs leaving a 5'-monophosphate group. Next, a synthetic RNA adaptor is ligated to the RNA population, only molecules which have a 5-phosphate (uncapped, full length mRNAs) will accept the adaptor. Reverse trascriptase reactions using an oligodT primer, and nested PCR (using one adaptor primer (located in the 5'

synthetic adaptor) and one primer for the gene) are then used to amplify the desired transcript.

In a preferred embodiment of this invention, the upper strand or lower strand primer may be also biotinylated or labeled at the 5' end with one of a) free amino group, b) thiol, c) carboxylic acid and d) another group not found in DNA that can react to form a strong bond to a known partner as an insoluble medium. These can then be used to immobilize the labeled strand after amplification. The immobilized DNA can be either single or double-stranded.

After amplification (using *e.g.*, RT CapExtension or RACE), the DNAs of this invention are rendered single-stranded. For example, the strands can be separated by using a biotinylated primer, capturing the biotinylated product on streptavidin beads, denaturing the DNA, and washing away the complementary strand. Depending on which end of the captured DNA is wanted, one will choose to immobilize either the upper (sense) strand or the lower (antisense) strand.

To prepare the single-stranded amplified DNAs for cloning into genetic packages so as to effect display of, or for expression of, the peptides, polypeptides or proteins encoded, at least in part, by those DNAs, they must be manipulated to provide ends suitable for cloning and display and/or expression. In particular, any 5' untranslated regions and mammalian signal sequences must be removed and replaced, in frame, by a suitable signal sequence that functions in the display or expression host. Additionally, parts of the variable domains (in antibody genes) may be removed and replaced by synthetic segments containing synthetic diversity. The diversity of other gene families may likewise be expanded with synthetic diversity.

According to the methods of this invention, there are two ways to manipulate the single-stranded DNAs for display and/or expression. The first method comprises the steps of:

(i) contacting the nucleic acid with a single-stranded oligonucleotide, the oligonucleotide being functionally complementary to the nucleic acid in the region in which cleavage is desired and including a sequence that with its complement in the nucleic acid forms a restriction endonuclease recognition site that on restriction results in cleavage of the nucleic acid at the desired location; and

(ii) cleaving the nucleic acid solely at the recognition site formed by the complementation of the nucleic acid and the oligonucleotide;

the contacting and the cleaving steps being performed at a temperature sufficient to maintain the nucleic acid in substantially single-stranded form, the oligonucleotide being functionally complementary to the nucleic acid over a large enough region to allow the two strands to associate such that cleavage may occur at the chosen temperature and at the desired location, and the cleavage being carried out using a restriction endonuclease that is active at the chosen temperature.

In this first method, short oligonucleotides are annealed to the single-stranded DNA so that restriction endonuclease recognition sites formed within the now locally double-stranded regions of the DNA can be cleaved. In particular, a recognition site

that occurs at the same position in a substantial fraction of the single-stranded DNAs is identical.

For antibody genes, this can be done using a catalog of germline sequences. See, *e.g.*, "http://www.mrc-cpe.cam.ac.uk/imt-doc/restricted/ok.html." Updates can be obtained from this site under the heading "Amino acid and nucleotide sequence alignments." For other families, similar comparisons exist and may be used to select appropriate regions for cleavage and to maintain diversity.

For example, Table 1 depicts the DNA sequences of the FR3 regions of the 51 known human VH germline genes. In this region, the genes contain restriction endonuclease recognition sites shown in Table 2. Restriction endonucleases that cleave a large fraction of germline genes at the same site are preferred over endonucleases that cut at a variety of sites. Furthermore, it is preferred that there be only one site for the restriction endonucleases within the region to which the short oligonucleotide binds on the single-stranded DNA, *e.g.*, about 10 bases on either side of the restriction endonuclease recognition site.

An enzyme that cleaves downstream in FR3 is also more preferable because it captures fewer mutations in the framework. This may be advantageous is some cases. However, it is well known that framework mutations exist and confer and enhance antibody binding. The present invention, by choice of appropriate restriction site, allows all or part of FR3 diversity to be captured. Hence, the method also allows extensive diversity to be captured.

Finally, in the methods of this invention restriction endonucleases that are active between about 37°C and about 75°C are used. Preferably, restriction

endonucleases that are active between about 45°C and about 75°C may be used. More preferably, enzymes that are active above 50°C, and most preferably active about 55°C, are used. Such temperatures maintain the nucleic acid sequence to be cleaved in substantially single-stranded form.

Enzymes shown in Table 2 that cut many of the heavy chain FR3 germline genes at a single position include: *Mae*III(24@4), *Tsp*45I(21@4), *Hph*I(44@5), *Bsa*JI(23@65), *Alu*I(23@47), *Blp*I(21@48), *Dde*I(29@58), *Bgl*II(10@61), *Msl*I(44@72), *Bsi*EI(23@74), *Eae*I(23@74), *Eag*I(23@74), *Hae*III(25@75), *Bst*4CI(51@86), *Hpy*CH4III(51@86), *Hin*fI(38@2), *Mly*I(18@2), *Ple*I(18@2), *Mnl*I(31@67), *Hpy*CH4V(21@44), *Bsm*AI(16@11), *Bpm*I(19@12), *Xmn*I(12@30), and *Sac*I(11@51). (The notation used means, for example, that *Bsm*AI cuts 16 of the FR3 germline genes with a restriction endonuclease recognition site beginning at base 11 of FR3.)

For cleavage of human heavy chains in FR3, the preferred restriction endonucleases are: *Bst*4CI (or *Taa*I or *Hpy*CH4III), *Blp*I, *Hpy*CH4V, and *Msl*I. Because ACNGT (the restriction endonuclease recognition site for *Bst*4CI, *Taa*I, and *Hpy*CH4III) is found at a consistent site in all the human FR3 germline genes, one of those enzymes is the most preferred for capture of heavy chain CDR3 diversity. *Blp*I and *Hpy*CH4V are complementary. *Blp*I cuts most members of the VH1 and VH4 families while *Hpy*CH4V cuts most members of the VH3, VH5, VH6, and VH7 families. Neither enzyme cuts VH2s, but this is a very small family, containing only three members. Thus, these enzymes may also be used in preferred embodiments of the methods of this invention.

The restriction endonucleases *Hpy*CH4III, *Bst*4CI, and *Taa*I all recognize 5'-ACnGT-3' and cut upper strand DNA after n and lower strand DNA before the base complementary to n. This is the most preferred restriction endonuclease recognition site for this method on human heavy chains because it is found in all germline genes. Furthermore, the restriction endonuclease recognition region (ACnGT) matches the second and third bases of a tyrosine codon (ta<u>y</u>) and the following cysteine codon (tg<u>y</u>) as shown in Table 3. These codons are highly conserved, especially the cysteine in mature antibody genes.

Table 4 E shows the distinct oligonucleotides of length 22 (except the last one which is of length 20) bases. Table 5 C shows the analysis of 1617 actual heavy chain antibody genes. Of these, 1511 have the site and match one of the candidate oligonucleotides to within 4 mismatches. Eight oligonucleotides account for most of the matches and are given in Table 4 F.1. The 8 oligonucleotides are very similar so that it is likely that satisfactory cleavage will be achieved with only one oligonucleotide (such as H43.77.97.1-02#1) by adjusting temperature, pH, salinity, and the like. One or two oligonucleotides may likewise suffice whenever the germline gene sequences differ very little and especially if they differ very little close to the restriction endonuclease recognition region to be cleaved. Table 5 D shows a repeat analysis of 1617 actual heavy chain antibody genes using only the 8 chosen oligonucleotides. This shows that 1463 of the sequences match at least one of the oligonucleotides to within 4 mismatches and have the site as expected.

Only 7 sequences have a second *Hpy*CH4III restriction endonuclease recognition region in this region.

Another illustration of choosing an appropriate restriction endonuclease recognition site involves cleavage in FR1 of human heavy chains. Cleavage in FR1 allows capture of the entire CDR diversity of the heavy chain.

The germline genes for human heavy chain FR1 are shown in Table 6. Table 7 shows the restriction endonuclease recognition sites found in human germline genes FR1s. The preferred sites are *Bsg*I(GTGCAG;39@4), *Bso*FI(GCngc;43@6,11@9,2@3,1@12), *Tse*I(Gcwgc;43@6,11@9,2@3,1@12), *Msp*A1I(CMGckg;46@7,2@1), *Pvu*II(CAGctg;46@7,2@1), *Alu*I(AGct;48@82@2), *Dde*I(Ctnag;22@52,9@48), *Hph*I(tcacc;22@80), *Bss*KI(Nccngg;35@39,2@40), *Bsa*JI(Ccnngg;32@40,2@41), *Bst*NI(CCwgg;33@40), *Scr*FI(CCngg;35@40,2@41), *Eco*0109I(RGgnccy;22@46, 11@43), *Sau*96I(Ggncc;23@47,11@44), AvaII(Ggwcc;23@47,4@44), *Ppu*MI(RGgwccy;22@46,4@43), *Bsm*FI(gtccc;20@48), *Hinf*I(Gantc;34@16,21@56,21@77), *Tfi*I(21@77), *Mly*I(GAGTC;34@16), *Mly*I(gactc;21@56), and *Alw*NI(CAGnnnctg;22@68). The more preferred sites are *MspAI* and *PvuII*. *Msp*AI and *Pvu*II have 46 sites at 7-12 and 2 at 1-6. To avoid cleavage at both sites, oligonucleotides are used that do not fully cover the site at 1-6. Thus, the DNA will not be cleaved at that site. We have shown that DNA that extends 3, 4, or 5 bases beyond a *Pvu*II-site can be cleaved efficiently.

Another illustration of choosing an appropriate restriction endonuclease recognition site involves cleavage in FR1 of human kappa light chains. Table 8 shows the human kappa FR1 germline genes and

Table 9 shows restriction endonuclease recognition sites that are found in a substantial number of human kappa FR1 germline genes at consistent locations. Of the restriction endonuclease recognition sites listed, *Bsm*AI and *Pfl*FI are the most preferred enzymes. *Bsm*AI sites are found at base 18 in 35 of 40 germline genes. *Pfl*FI sites are found in 35 of 40 germline genes at base 12.

Another example of choosing an appropriate restriction endonuclease recognition site involves cleavage in FR1 of the human lambda light chain. Table 10 shows the 31 known human lambda FR1 germline gene sequences. Table 11 shows restriction endonuclease recognition sites found in human lambda FR1 germline genes. *Hin*fI and *Dde*I are the most preferred restriction endonucleases for cutting human lambda chains in FR1.

After the appropriate site or sites for cleavage are chosen, one or more short oligonucleotides are prepared so as to functionally complement, alone or in combination, the chosen recognition site. The oligonucleotides also include sequences that flank the recognition site in the majority of the amplified genes. This flanking region allows the sequence to anneal to the single-stranded DNA sufficiently to allow cleavage by the restriction endonuclease specific for the site chosen.

The actual length and sequence of the oligonucleotide depends on the recognition site and the conditions to be used for contacting and cleavage. The length must be sufficient so that the oligonucleotide is functionally complementary to the single-stranded DNA over a large enough region to allow the two strands

to associate such that cleavage may occur at the chosen temperature and at the desired location.

Typically, the oligonucleotides of this preferred method of the invention are about 17 to about 30 nucleotides in length. Below about 17 bases, annealing is too weak and above 30 bases there can be a loss of specificity. A preferred length is 18 to 24 bases.

Oligonucleotides of this length need not be identical complements of the germline genes. Rather, a few mismatches taken may be tolerated. Preferably, however, no more than 1-3 mismatches are allowed. Such mismatches do not adversely affect annealing of the oligonucleotide to the single-stranded DNA. Hence, the two DNAs are said to be functionally complementary.

The second method to manipulate the single-stranded DNAs of this invention for display and/or expression comprises the steps of:

(i) contacting the nucleic acid with a partially double-stranded oligonucleotide, the single-stranded region of the oligonucleotide being functionally complementary to the nucleic acid in the region in which cleavage is desired, and the double-stranded region of the oligonucleotide having a restriction endonuclease recognition site; and

(ii) cleaving the nucleic acid solely at the cleavage site formed by the complementation of the nucleic acid and the single-stranded region of the oligonucleotide;

the contacting and the cleaving steps being performed at a temperature sufficient to maintain the nucleic acid in substantially single-stranded form, the oligonucleotide being functionally complementary to the nucleic acid over a large enough region to allow the two strands to associate such that cleavage may occur at the chosen temperature and at the desired location, and the cleavage being carried out using a restriction endonuclease that is active at the chosen temperature.

As explained above, the cleavage site may be formed by the single-stranded portion of the partially double-stranded oligonucleotide duplexing with the single-stranded DNA, the cleavage site may be carried in the double-stranded portion of the partially double-stranded oligonucleotide, or the cleavage site may be introduced by the amplification primer used to amplify the single-stranded DNA-partially double-stranded oligonucleotide combination. In this embodiment, the first is preferred. And, the restriction endonuclease recognition site may be located in either the double-stranded portion of the oligonucleotide or introduced by the amplification primer, which is complementary to that double-stranded region, as used to amplify the combination.

Preferably, the restriction endonuclease site is that of a Type II-S restriction endonuclease, whose cleavage site is located at a known distance from its recognition site.

This second method, preferably, employs Universal Restriction Endonucleases ("URE"). UREs are partially double-stranded oligonucleotides. The single-stranded portion or overlap of the URE consists of a DNA adapter that is functionally complementary to the sequence to be cleaved in the single-stranded DNA.

The double-stranded portion consists of a restriction endonuclease recognition site, preferably type II-S.

The URE method of this invention is specific and precise and can tolerate some (*e.g.*, 1-3) mismatches in the complementary regions, *i.e.*, it is functionally complementary to that region. Further, conditions under which the URE is used can be adjusted so that most of the genes that are amplified can be cut, reducing bias in the library produced from those genes.

The sequence of the single-stranded DNA adapter or overlap portion of the URE typically consists of about 14-22 bases. However, longer or shorter adapters may be used. The size depends on the ability of the adapter to associate with its functional complement in the single-stranded DNA and the temperature used for contacting the URE and the single-stranded DNA at the temperature used for cleaving the DNA with the restriction enzyme. The adapter must be functionally complementary to the single-stranded DNA over a large enough region to allow the two strands to associate such that the cleavage may occur at the chosen temperature and at the desired location. We prefer singe-stranded or overlap portions of 14-17 bases in length, and more preferably 18-20 bases in length.

The site chosen for cleavage using the URE is preferably one that is substantially conserved in the family of amplified DNAs. As compared to the first cleavage method of this invention, these sites do not need to be endonuclease recognition sites. However, like the first method, the sites chosen can be synthetic rather than existing in the native DNA. Such sites may be chosen by references to the sequences of

known antibodies or other families of genes. For example, the sequences of many germline genes are reported at http://www.mrc-cpe.cam.ac.uk/imt-doc/restricted/ok.html. For example, one preferred site occurs near the end of FR3 -- codon 89 through the second base of codon 93. CDR3 begins at codon 95.

The sequences of 79 human heavy-chain genes are also available at http://www.ncbi.nlm.nih.gov/entre2/nucleotide.html. This site can be used to identify appropriate sequences for URE cleavage according to the methods of this invention. *See, e.g.*, Table 12B.

Most preferably, one or more sequences are identified using these sites or other available sequence information. These sequences together are present in a substantial fraction of the amplified DNAs. For example, multiple sequences could be used to allow for known diversity in germline genes or for frequent somatic mutations. Synthetic degenerate sequences could also be used. Preferably, a sequence(s) that occurs in at least 65% of genes examined with no more than 2-3 mismatches is chosen

URE single-stranded adapters or overlaps are then made to be complementary to the chosen regions. Conditions for using the UREs are determined empirically. These conditions should allow cleavage of DNA that contains the functionally complementary sequences with no more than 2 or 3 mismatches but that do not allow cleavage of DNA lacking such sequences.

As described above, the double-stranded portion of the URE includes an endonuclease recognition site, preferably a Type II-S recognition site. Any enzyme that is active at a temperature necessary to maintain the single-stranded DNA substantially in that

form and to allow the single-stranded DNA adapter portion of the URE to anneal long enough to the single-stranded DNA to permit cleavage at the desired site may be used.

The preferred Type II-S enzymes for use in the URE methods of this invention provide asymmetrical cleavage of the single-stranded DNA. Among these are the enzymes listed in Table 13. The most preferred Type II-S enzyme is *FokI*.

When the preferred *FokI* containing URE is used, several conditions are preferably used to effect cleavage:

1) Excess of the URE over target DNA should be present to activate the enzyme. URE present only in equimolar amounts to the target DNA would yield poor cleavage of ssDNA because the amount of active enzyme available would be limiting.
2) An activator may be used to activate part of the *Fok*I enzyme to dimerize without causing cleavage. Examples of appropriate activators are shown in Table 14.
3) The cleavage reaction is performed at a temperature between 45°-75°C, preferably above 50°C and most preferably above 55°C.

The UREs used in the prior art contained a 14-base single-stranded segment, a 10-base stem (containing a *FokI* site), followed by the palindrome of the 10-base stem. While such UREs may be used in the methods of this invention, the preferred UREs of this invention also include a segment of three to eight bases (a loop) between the *FokI* restriction

endonuclease recognition site containing segments. In the preferred embodiment, the stem (containing the *FokI* site) and its palindrome are also longer than 10 bases. Preferably, they are 10-14 bases in length. Examples of these "lollipop" URE adapters are shown in Table 15.

One example of using a URE to cleave an single-stranded DNA involves the FR3 region of human heavy chain. Table 16 shows an analysis of 840 full-length mature human heavy chains with the URE recognition sequences shown. The vast majority (718/840=0.85) will be recognized with 2 or fewer mismatches using five UREs (VHS881-1.1, VHS881-1.2, VHS881-2.1, VHS881-4.1, and VHS881-9.1). Each has a 20-base adaptor sequence to complement the germline gene, a ten-base stem segment containing a *Fok*I site, a five base loop, and the reverse complement of the first stem segment. Annealing those adapters, alone or in combination, to single-stranded antisense heavy chain DNA and treating with *Fok*I in the presence of, *e.g.*, the activator FOKIact, will lead to cleavage of the antisense strand at the position indicated.

Another example of using a URE(s) to cleave a single-stranded DNA involves the FR1 region of the human Kappa light chains. Table 17 shows an analysis of 182 full-length human kappa chains for matching by the four 19-base probe sequences shown. Ninety-six percent of the sequences match one of the probes with 2 or fewer mismatches. The URE adapters shown in Table 17 are for cleavage of the sense strand of kappa chains. Thus, the adaptor sequences are the reverse complement of the germline gene sequences. The URE consists of a ten-base stem, a five base loop, the reverse complement of the stem and the complementation

sequence. The loop shown here is TTGTT, but other sequences could be used. Its function is to interrupt the palindrome of the stems so that formation of a lollypop monomer is favored over dimerization. Table 17 also shows where the sense strand is cleaved.

Another example of using a URE to cleave a single-stranded DNA involves the human lambda light chain. Table 18 shows analysis of 128 human lambda light chains for matching the four 19-base probes shown. With three or fewer mismatches, 88 of 128 (69%) of the chains match one of the probes. Table 18 also shows URE adapters corresponding to these probes. Annealing these adapters to upper-strand ssDNA of lambda chains and treatment with *Fok*I in the presence of FOKIact at a temperature at or above 45°C will lead to specific and precise cleavage of the chains.

The conditions under which the short oligonucleotide sequences of the first method and the UREs of the second method are contacted with the single-stranded DNAs may be empirically determined. The conditions must be such that the single-stranded DNA remains in substantially single-stranded form. More particularly, the conditions must be such that the single-stranded DNA does not form loops that may interfere with its association with the oligonucleotide sequence or the URE or that may themselves provide sites for cleavage by the chosen restriction endonuclease.

The effectiveness and specificity of short oligonucleotides (first method) and UREs (second method) can be adjusted by controlling the concentrations of the URE adapters/oligonucleotides and substrate DNA, the temperature, the pH, the concentration of metal ions, the ionic strength, the

concentration of chaotropes (such as urea and formamide), the concentration of the restriction endonuclease(*e.g.*, *Fok*I), and the time of the digestion. These conditions can be optimized with synthetic oligonucleotides having: 1) target germline gene sequences, 2) mutated target gene sequences, or 3) somewhat related non-target sequences. The goal is to cleave most of the target sequences and minimal amounts of non-targets.

In accordance with this invention, the single-stranded DNA is maintained in substantially that form using a temperature between about 37°C and about 75°C. Preferably, a temperature between about 45°C and about 75°C is used. More preferably, a temperature between 50°C and 60°C, most preferably between 55°C and 60°C, is used. These temperatures are employed both when contacting the DNA with the oligonucleotide or URE and when cleaving the DNA using the methods of this invention.

The two cleavage methods of this invention have several advantages. The first method allows the individual members of the family of single-stranded DNAs to be cleaved preferentially at one substantially conserved endonuclease recognition site. The method also does not require an endonuclease recognition site to be built into the reverse transcription or amplification primers. Any native or synthetic site in the family can be used.

The second method has both of these advantages. In addition, the preferred URE method allows the single-stranded DNAs to be cleaved at positions where no endonuclease recognition site naturally occurs or has been synthetically constructed.

Most importantly, both cleavage methods permit the use of 5' and 3' primers so as to maximize diversity and then cleavage to remove unwanted or deleterious sequences before cloning, display and/or expression.

After cleavage of the amplified DNAs using one of the methods of this invention, the DNA is prepared for cloning, display and/or expression. This is done by using a partially duplexed synthetic DNA adapter, whose terminal sequence is based on the specific cleavage site at which the amplified DNA has been cleaved.

The synthetic DNA is designed such that when it is ligated to the cleaved single-stranded DNA in proper reading frame so that the desired peptide, polypeptide or protein can be displayed on the surface of the genetic package and/or expressed. Preferably, the double-stranded portion of the adapter comprises the sequence of several codons that encode the amino acid sequence characteristic of the family of peptides, polypeptides or proteins up to the cleavage site. For human heavy chains, the amino acids of the 3-23 framework are preferably used to provide the sequences required for expression of the cleaved DNA.

Preferably, the double-stranded portion of the adapter is about 12 to 100 bases in length. More preferably, about 20 to 100 bases are used. The double-standard region of the adapter also preferably contains at least one endonuclease recognition site useful for cloning the DNA into a suitable display and/or expression vector (or a recipient vector used to archive the diversity). This endonuclease restriction site may be native to the germline gene sequences used to extend the DNA sequence. It may be also constructed

using degenerate sequences to the native germline gene sequences. Or, it may be wholly synthetic.

The single-stranded portion of the adapter is complementary to the region of the cleavage in the single-stranded DNA. The overlap can be from about 2 bases up to about 15 bases. The longer the overlap, the more efficient the ligation is likely to be. A preferred length for the overlap is 7 to 10. This allows some mismatches in the region so that diversity in this region may be captured.

The single-stranded region or overlap of the partially duplexed adapter is advantageous because it allows DNA cleaved at the chosen site, but not other fragments to be captured. Such fragments would contaminate the library with genes encoding sequences that will not fold into proper antibodies and are likely to be non-specifically sticky.

One illustration of the use of a partially duplexed adaptor in the methods of this invention involves ligating such adaptor to a human FR3 region that has been cleaved, as described above, at 5'-ACnGT-3' using HpyCH4III, Bst4CI or TaaI.

Table 4 F.2 shows the bottom strand of the double-stranded portion of the adaptor for ligation to the cleaved bottom-strand DNA. Since the HpyCH4III-Site is so far to the right (as shown in Table 3), a sequence that includes the *Afl*II-site as well as the *Xba*I site can be added. This bottom strand portion of the partially-duplexed adaptor, H43.XAExt, incorporates both *Xba*I and *Afl*II-sites. The top strand of the double-stranded portion of the adaptor has neither site (due to planned mismatches in the segments opposite the *Xba*I and *Afl*II-Sites of H43.XAExt), but

will anneal very tightly to H43.XAExt. H43AExt contains only the *Afl*II-site and is to be used with the top strands H43.ABr1 and H43.ABr2 (which have intentional alterations to destroy the *Afl*II-site).

After ligation, the desired, captured DNA can be PCR amplified again, if desired, using in the preferred embodiment a primer to the downstream constant region of the antibody gene and a primer to part of the double-standard region of the adapter. The primers may also carry restriction endonuclease sites for use in cloning the amplified DNA.

After ligation, and perhaps amplification, of the partially double-stranded adapter to the single-stranded amplified DNA, the composite DNA is cleaved at chosen 5' and 3' endonuclease recognition sites.

The cleavage sites useful for cloning depend on the phage or phagemid or other vectors into which the cassette will be inserted and the available sites in the antibody genes. Table 19 provides restriction endonuclease data for 75 human light chains. Table 20 shows corresponding data for 79 human heavy chains. In each Table, the endonucleases are ordered by increasing frequency of cutting. In these Tables, Nch is the number of chains cut by the enzyme and Ns is the number of sites (some chains have more than one site).

From this analysis, *Sfi*I, *Not*I, *Afl*II, *Apa*LI, and *Asc*I are very suitable. *Sfi*I and *Not*I are preferably used in pCES1 to insert the heavy-chain display segment. *Apa*LI and *Asc*I are preferably used in pCES1 to insert the light-chain display segment.

*Bst*EII-sites occur in 97% of germ-line JH genes. In rearranged V genes, only 54/79 (68%) of heavy-chain genes contain a *Bst*EII-Site and 7/61 of

these contain two sites. Thus, 47/79 (59%) contain a single *Bst*EII-Site. An alternative to using *Bst*EII is to cleave *via* UREs at the end of JH and ligate to a synthetic oligonucleotide that encodes part of CH1.

One example of preparing a family of DNA sequences using the methods of this invention involves capturing human CDR 3 diversity. As described above, mRNAs from various autoimmune patients are reverse transcribed into lower strand cDNA. After the top strand RNA is degraded, the lower strand is immobilized and a short oligonucleotide used to cleave the cDNA upstream of CDR3. A partially duplexed synthetic DNA adapter is then annealed to the DNA and the DNA is amplified using a primer to the adapter and a primer to the constant region (after FR4). The DNA is then cleaved using *BstEII* (in FR4) and a restriction endonuclease appropriate to the partially double-stranded adapter (*e.g.*, *XbaI* and *AflII* (in FR3)). The DNA is then ligated into a synthetic VH skeleton such as 3-23.

One example of preparing a single-stranded DNA that was cleaved using the URE method involves the human Kappa chain. The cleavage site in the sense strand of this chain is depicted in Table 17. The oligonucleotide kapextURE is annealed to the oligonucleotides (kaBR01UR, kaBR02UR, kaBR03UR, and kaBR04UR) to form a partially duplex DNA. This DNA is then ligated to the cleaved soluble kappa chains. The ligation product is then amplified using primers kapextUREPCR and CKForeAsc (which inserts a AscI site after the end of C kappa). This product is then cleaved with *Apa*LI and *Asc*I and ligated to similarly cut recipient vector.

Another example involves the cleavage of lambda light chains, illustrated in Table 18. After cleavage, an extender (ON_LamEx133) and four bridge oligonucleotides (ON_LamB1-133, ON_LamB2-133, ON_LamB3-133, and ON_LamB4-133) are annealed to form a partially duplex DNA. That DNA is ligated to the cleaved lambda-chain sense strands. After ligation, the DNA is amplified with ON_Lam133PCR and a forward primer specific to the lambda constant domain, such as CL2ForeAsc or CL7ForeAsc (Table 130).

In human heavy chains, one can cleave almost all genes in FR4 (downstream, *i.e.*, toward the 3' end of the sense strand, of CDR3) at a *Bst*EII-Site that occurs at a constant position in a very large fraction of human heavy-chain V genes. One then needs a site in FR3, if only CDR3 diversity is to be captured, in FR2, if CDR2 and CDR3 diversity is wanted, or in FR1, if all the CDR diversity is wanted. These sites are preferably inserted as part of the partially double-stranded adaptor.

The preferred process of this invention is to provide recipient vectors (*e.g.*, for display and/or expression) having sites that allow cloning of either light or heavy chains. Such vectors are well known and widely used in the art. A preferred phage display vector in accordance with this invention is phage MALIA3. This displays in gene III. The sequence of the phage MALIA3 is shown in Table 21A (annotated) and Table 21B (condensed).

The DNA encoding the selected regions of the light or heavy chains can be transferred to the vectors using endonucleases that cut either light or heavy chains only very rarely. For example, light chains may

be captured with *Apa*LI and *Asc*I. Heavy-chain genes are preferably cloned into a recipient vector having *Sfi*I, *Nco*I, *Xba*I, *Afl*II, *Bst*EII, *Apa*I, and *Not*I sites. The light chains are preferably moved into the library as *Apa*LI-*Asc*I fragments. The heavy chains are preferably moved into the library as *Sfi*I-*Not*I fragments.

Most preferably, the display is had on the surface of a derivative of M13 phage. The most preferred vector contains all the genes of M13, an antibiotic resistance gene, and the display cassette. The preferred vector is provided with restriction sites that allow introduction and excision of members of the diverse family of genes, as cassettes. The preferred vector is stable against rearrangement under the growth conditions used to amplify phage.

In another embodiment of this invention, the diversity captured by the methods of the present invention may be displayed and/or expressed in a phagemid vector (*e.g.*, pCES1) that displays and/or expresses the peptide, polypeptide or protein. Such vectors may also be used to store the diversity for subsequent display and/or expression using other vectors or phage.

In another embodiment of this invention, the diversity captured by the methods of the present invention may be displayed and/or expressed in a yeast vector.

In another embodiment, the mode of display may be through a short linker to anchor domains -- one possible anchor comprising the final portion of M13 III ("IIIstump") and a second possible anchor being the full length III mature protein.

The IIIstump fragment contains enough of M13

III to assemble into phage but not the domains involved in mediating infectivity. Because the w.t. III proteins are present the phage is unlikely to delete the antibody genes and phage that do delete these segments receive only a very small growth advantage. For each of the anchor domains, the DNA encodes the w.t. AA sequence, but differs from the w.t. DNA sequence to a very high extent. This will greatly reduce the potential for homologous recombination between the anchor and the w.t. gene that is also present (see Example 6).

Most preferably, the present invention uses a complete phage carrying an antibiotic-resistance gene (such as an ampicillin-resistance gene) and the display cassette. Because the w.t. *iii* and possibly *viii* genes are present, the w.t. proteins are also present. The display cassette is transcribed from a regulatable promoter (*e.g.*, P_{Lacz}). Use of a regulatable promoter allows control of the ratio of the fusion display gene to the corresponding w.t. coat protein. This ratio determines the average number of copies of the display fusion per phage (or phagemid) particle.

Another aspect of the invention is a method of displaying peptides, polypeptides or proteins (and particularly Fabs) on filamentous phage. In the most preferred embodiment this method displays FABs and comprises:

a) obtaining a cassette capturing a diversity of segments of DNA encoding the elements:

P_{reg}::RBS1::SS1::VL::CL::stop::RBS2::SS2::VH::CH1::linker::anchor::stop::,

where P_{reg} is a regulatable promoter, RBS1 is a first

ribosome binding site, SS1 is a signal sequence operable in the host strain, VL is a member of a diverse set of light-chain variable regions, CL is a light-chain constant region, stop is one or more stop codons, RBS2 is a second ribosome binding site, SS2 is a second signal sequence operable in the host strain, VH is a member of a diverse set of heavy-chain variable regions, CH1 is an antibody heavy-chain first constant domain, linker is a sequence of amino acids of one to about 50 residues, anchor is a protein that will assemble into the filamentous phage particle and stop is a second example of one or more stop codons; and

b) positioning that cassette within the phage genome to maximize the viability of the phage and to minimize the potential for deletion of the cassette or parts thereof.

The DNA encoding the anchor protein in the above preferred cassette should be designed to encode the same (or a closely related) amino acid sequence as is found in one of the coat proteins of the phage, but with a distinct DNA sequence. This is to prevent unwanted homologous recombination with the w.t. gene. In addition, the cassette should be placed in the intergenic region. The positioning and orientation of the display cassette can influence the behavior of the phage.

In one embodiment of the invention, a transcription terminator may be placed after the second stop of the display cassette above (*e.g.*, Trp). This will reduce interaction between the display cassette and other genes in the phage antibody display vector.

In another embodiment of the methods of this invention, the phage or phagemid can display and/or

express proteins other than Fab, by replacing the Fab portions indicated above, with other protein genes.

Various hosts can be used the display and/or expression aspect of this invention. Such hosts are well known in the art. In the preferred embodiment, where Fabs are being displayed and/or expressed, the preferred host should grow at 30°C and be $RecA^-$ (to reduce unwanted genetic recombination) and $EndA^-$ (to make recovery of RF DNA easier). It is also preferred that the host strain be easily transformed by electroporation.

XL1-Blue MRF' satisfies most of these preferences, but does not grow well at 30°C. XL1-Blue MRF' does grow slowly at 38°C and thus is an acceptable host. TG-1 is also an acceptable host although it is $RecA^+$ and $EndA^+$. XL1-Blue MRF' is more preferred for the intermediate host used to accumulate diversity prior to final construction of the library.

After display and/or expression, the libraries of this invention may be screened using well known and conventionally used techniques. The selected peptides, polypeptides or proteins may then be used to treat disease. Generally, the peptides, polypeptides or proteins for use in therapy or in pharmaceutical compositions are produced by isolating the DNA encoding the desired peptide, polypeptide or protein from the member of the library selected. That DNA is then used in conventional methods to produce the peptide, polypeptides or protein it encodes in appropriate host cells, preferably mammalian host cells, *e.g.*, CHO cells. After isolation, the peptide, polypeptide or protein is used alone or with pharmaceutically acceptable compositions in therapy to treat disease.

EXAMPLES

Example 1: RACE amplification of heavy and light chain antibody repertoires from autoimmune patients.

Total RNA was isolated from individual blood samples (50 ml) of 11 patients using a RNAzolTM kit (CINNA/Biotecx), as described by the manufacturer. The patients were diagnosed as follows:

1. SLE and phospholipid syndrome
2. limited systemic sclerosis
3. SLE and Sjogren syndrome
4. Limited Systemic sclerosis
5. Reumatoid Arthritis with active vasculitis
6. Limited systemic sclerosis and Sjogren Syndrome
7. Reumatoid Artritis and (not active) vasculitis
8. SLE and Sjogren syndrome
9. SLE
10. SLE and (active) glomerulonephritis
11. Polyarthritis/ Raynauds Phenomen

From these 11 samples of total RNA, Poly-A+ RNA was isolated using Promega PolyATtract® mRNA Isolation kit (Promega).

250 ng of each poly-A+ RNA sample was used to amplify antibody heavy and light chains with the GeneRAacerTM kit (Invitrogen cat no. L1500-01). A schematic overview of the RACE procedure is shown in FIG. 3.

Using the general protocol of the GeneRAacer™ kit, an RNA adaptor was ligated to the 5'end of all mRNAs. Next, a reverse transcriptase reaction was performed in the presence of oligo(dT15) specific

primer under conditions described by the manufacturer in the GeneRAacer™ kit.

1/5 of the cDNA from the reverse transcriptase reaction was used in a 20 ul PCR reaction. For amplification of the heavy chain IgM repertoire, a forward primer based on the CH1 chain of IgM [HuCmFOR] and a backward primer based on the ligated synthetic adaptor sequence [5'A] were used. (See Table 22)

For amplification of the kappa and lambda light chains, a forward primer that contains the 3' coding-end of the cDNA [HuCkFor and HuCLFor2+HuCLfor7] and a backward primer based on the ligated synthetic adapter sequence [5'A] was used (See Table 22). Specific amplification products after 30 cycles of primary PCR were obtained.

FIG. 4 shows the amplification products obtained after the primary PCR reaction from 4 different patient samples. 8 ul primary PCR product from 4 different patients was analyzed on a agarose gel [labeled 1,2, 3 and 4]. For the heavy chain, a product of approximately 950 nt is obtained while for the kappa and lambda light chains the product is approximately 850 nt. M1-2 are molecular weight markers.

PCR products were also analyzed by DNA sequencing [10 clones from the lambda, kappa or heavy chain repertoires]. All sequenced antibody genes recovered contained the full coding sequence as well as the 5' leader sequence and the V gene diversity was the expected diversity (compared to literature data).

50 ng of all samples from all 11 individual amplified samples were mixed for heavy, lambda light or kappa light chains and used in secondary PCR reactions.

In all secondary PCRs approximately 1 ng

template DNA from the primary PCR mixture was used in multiple 50 ul PCR reactions [25 cycles].

For the heavy chain, a nested biotinylated forward primer [HuCm-Nested] was used, and a nested 5'end backward primer located in the synthetic adapter-sequence [5'NA] was used. The 5'end lower-strand of the heavy chain was biotinylated.

For the light chains, a 5'end biotinylated nested primer in the synthetic adapter was used [5'NA] in combination with a 3'end primer in the constant region of Ckappa and Clambda, extended with a sequence coding for the AscI restriction site [kappa: HuCkForAscI, Lambda: HuCL2-FOR-ASC + HuCL7-FOR-ASC]. [5'end Top strand DNA was biotinylated]. After gel-analysis the secondary PCR products were pooled and purified with Promega Wizzard PCR cleanup. Approximately 25 ug biotinylated heavy chain, lambda and kappa light chain DNA was isolated from the 11 patients.

Example 2: Capturing kappa chains with BsmAI.

A repertoire of human-kappa chain mRNAs was prepared using the RACE method of Example 1 from a collection of patients having various autoimmune diseases.

This Example followed the protocol of Example 1. Approximately 2 micrograms (ug) of human kappa-chain (Igkappa) gene RACE material with biotin attached to 5'-end of upper strand was immobilized as in Example 1 on 200 microliters (µL) of Seradyn magnetic beads. The lower strand was removed by washing the DNA with 2 aliquots 200 µL of 0.1 M NaOH (pH 13) for 3 minutes for the first aliquot followed by 30 seconds for the second

aliquot. The beads were neutralized with 200 µL of 10 mM Tris (pH 7.5) 100 mM NaCl. The short oligonucleotides shown in Table 23 were added in 40 fold molar excess in 100 µL of NEB buffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol pH 7.9) to the dry beads. The mixture was incubated at 95°C for 5 minutes then cooled down to 55°C over 30 minutes. Excess oligonucleotide was washed away with 2 washes of NEB buffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol pH 7.9). Ten units of BsmAI (NEB) were added in NEB buffer 3 and incubated for 1 h at 55°C. The cleaved downstream DNA was collected and purified over a Qiagen PCR purification column (FIGs. 5 and 6).

FIG. 5 shows an analysis of digested kappa single-stranded DNA. Approximately 151.5 pmol of adapter was annealed to 3.79 pmol of immobilized kappa single-stranded DNA followed by digestion with 15 U of *BsmAI*. The supernatant containing the desired DNA was removed and analyzed by 5% polyacrylamide gel along with the remaining beads which contained uncleaved full length kappa DNA. 189 pmol of cleaved single-stranded DNA was purified for further analysis. Five percent of the original full length ssDNA remained on the beads.

FIG. 6 shows an analysis of the extender - cleaved kappa ligation. 180 pmol of pre-annealed bridge/extender was ligated to 1.8 pmol of *BsmAI* digested single-stranded DNA. The ligated DNA was purified by Qiagen PCR purification column and analyzed on a 5% polyacrylamide gel. Results indicated that the ligation of extender to single-stranded DNA was 95% efficient.

A partially double-stranded adaptor was prepared using the oligonucleotide shown in Table 23.

The adaptor was added to the single-stranded DNA in 100 fold molar excess along with 1000 units of T4 DNA ligase and incubated overnight at 16°C. The excess oligonucleotide was removed with a Qiagen PCR purification column. The ligated material was amplified by PCR using the primers kapPCRt1 and kapfor shown in Table 23 for 10 cycles with the program shown in Table 24.

The soluble PCR product was run on a gel and showed a band of approximately 700 n, as expected (FIGs. 7 and 8). The DNA was cleaved with enzymes *Apa*LI and *Asc*I, gel purified, and ligated to similarly cleaved vector pCES1.

FIG. 7 shows an analysis of the PCR product from the extender-kappa amplification. Ligated extender-kappa single-stranded DNA was amplified with primers specific to the extender and to the constant region of the light chain. Two different template concentrations, 10 ng versus 50 ng, were used as template and 13 cycles were used to generate approximately 1.5 ug of dsDNA as shown by 0.8% agarose gel analysis.

FIG. 8 shows an analysis of the purified PCR product from the extender-kappa amplification. Approximately 5 ug of PCR amplified extender-kappa double-stranded DNA was run out on a 0.8% agarose gel, cut out, and extracted with a GFX gel purification column. By gel analysis, 3.5 ug of double-stranded DNA was prepared.

The assay for capturing kappa chains with BsmA1 was repeated and produced similar results. FIG 9A shows the DNA after it was cleaved and collected and purified over a Qiagen PCR purification column. FIG. 9B shows the partially double-stranded adaptor

ligated to the single-stranded DNA. This ligated material was then amplified (FIG. 9C). The gel showed a band of approximately 700 n.

Table 25 shows the DNA sequence of a kappa light chain captured by this procedure. Table 26 shows a second sequence captured by this procedure. The closest bridge sequence was complementary to the sequence 5'-agccacc-3', but the sequence captured reads 5'-Tgccacc-3', showing that some mismatch in the overlapped region is tolerated.

Example 3: Construction of Synthetic CDR1 and CDR2 Diversity in V-3-23 VH Framework.

Synthetic diversity in Complementary Determinant Region (CDR) 1 and 2 was created in the 3-23 VH framework in a two step process: first, a vector containing the 3-23 VH framework was constructed; and then, a synthetic CDR 1 and 2 was assembled and cloned into this vector.

For construction of the 3-23 VH framework, 8 oligonucleotides and two PCR primers (long oligonucleotides - TOPFR1A, BOTFR1B, BOTFR2, BOTFR3, F06, BOTFR4, ON-vgC1, and ON-vgC2 and primers - SFPRMET and BOTPCRPRIM, shown in Table 27) that overlap were designed based on the Genebank sequence of 3-23 VH framework region. The design incorporated at least one useful restriction site in each framework region, as shown in Table 27. In Table 27, the segments that were synthesized are shown as bold, the overlapping regions are underscored, and the PCR priming regions at each end are underscored.

A mixture of these 8 oligos was combined at a final concentration of 2.5uM in a 20ul PCR reaction. The PCR mixture contained 200uM dNTPs, 2.5mM $MgCl_2$, 0.02U *Pfu Turbo*™ DNA Polymerase, 1U Qiagen HotStart Taq DNA Polymerase, and 1X Qiagen PCR buffer. The PCR program consisted of 10 cycles of 94°C for 30s, 55°C for 30s, and 72°C for 30s.

The assembled 3-23 VH DNA sequence was then amplified, using 2.5ul of a 10-fold dilution from the initial PCR in 100ul PCR reaction. The PCR reaction contained 200uM dNTPs, 2.5mM $MgCl_2$, 0.02U *Pfu Turbo*™ DNA Polymerase, 1U Qiagen HotStart Taq DNA Polymerase, 1X Qiagen PCR Buffer and 2 outside primers (SFPRMET and BOTPCRPRIM) at a concentration of 1uM. The PCR program consisted of 23 cycles at 94°C for 30s, 55°C for 30s, and 72°C for 60s. The 3-23 VH DNA sequence was digested and cloned into pCES1 (phagemid vector) using the *Sfi*I and *Bst*EII restriction endonuclease sites. All restriction enzymes mentioned herein were supplied by New England BioLabs, Beverly, MA and used as per the manufacturer's instructions.

Stuffer sequences (shown in Table 28 and Table 29) were introduced into pCES1 to replace CDR1/CDR2 sequences (900 bases between *Bsp*EI and *Xba*I RE sites) and CDR3 sequences (358 bases between *Afl*II and *Bst*EII) prior to cloning the CDR1/CDR2 diversity. This new vector was termed pCES5 and its sequence is given in Table 29.

Having stuffers in place of the CDRs avoids the risk that a parental sequence would be over-represented in the library. The stuffer sequences are fragments from the penicillase gene of *E. coli*. The CDR1-2 stuffer contains restriction sites for *Bgl*II,

*Bsu*36I, *Bcl*I, *Xcm*I, *Mlu*I, *Pvu*II, *Hpa*I, and *Hinc*II, the underscored sites being unique within the vector pCES5. The stuffer that replaces CDR3 contains the unique restriction endonuclease site *Rsr*II.

A schematic representation of the design for CDR1 and CDR2 synthetic diversity is shown FIG. 10. The design was based on the presence of mutations in DP47/3-23 and related germline genes. Diversity was designed to be introduced at the positions within CDR1 and CDR2 indicated by the numbers in FIG. 10. The diversity at each position was chosen to be one of the three following schemes: 1 = ADEFGHIKLMNPQRSTVWY; 2 = YRWVGS; 3 = PS, in which letters encode equimolar mixes of the indicated amino acids.

For the construction of the CDR1 and CDR2 diversity, 4 overlapping oligonucleotides (ON-vgC1, ON_Br12, ON_CD2Xba, and ON-vgC2, shown in Table 27 and Table 30) encoding CDR1/2, plus flanking regions, were designed. A mixture of these 4 oligos was combined at a final concentration of 2.5uM in a 40ul PCR reaction. Two of the 4 oligos contained variegated sequences positioned at the CDR1 and the CDR2. The PCR mixture contained 200uM dNTPs, 2.5U Pwo DNA Polymerase (Roche), and 1X Pwo PCR buffer with 2mM $MgSO_4$. The PCR program consisted of 10 cycles at 94°C for 30s, 60°C for 30s, and 72°C for 60s. This assembled CDR1/2 DNA sequence was amplified, using 2.5ul of the mixture in 100ul PCR reaction. The PCR reaction contained 200uM dNTPs, 2.5U Pwo DNA Polymerase, 1X Pwo PCR Buffer with 2mM $MgSO_4$ and 2 outside primers at a concentration of 1uM. The PCR program consisted of 10 cycles at 94°C for 30s, 60°C for 30s, and 72°C for 60s. These variegated sequences were digested and cloned into the 3-23 VH framework in place of the CDR1/2 stuffer.

We obtained approximately 7 X 10^7 independent transformants. CDR3 diversity either from donor populations or from synthetic DNA can be cloned into the vector containing synthetic CDR1 and CDR 2 diversity.

A schematic representation of this procedure is shown in FIG. 11. A sequence encoding the FR-regions of the human V3-23 gene segment and CDR regions with synthetic diversity was made by oligonucleotide assembly and cloning via *BspE1* and *Xbal* sites into a vector that complements the FR1 and FR3 regions. Into this library of synthetic VH segments, the complementary VH-CDR3 sequence (top right) was cloned via Xbal an BstEll sites. The resulting cloned CH genes contain a combination of designed synthetic diversity and natural diversity (see FIG. 11).

Example 4: Cleavage and ligation of the lambda light chains with HinfI.

A schematic of the cleavage and ligation of antibody light chains is shown in FIGs. 12A and 12B. Approximately 2 ug of biotinylated human Lambda DNA prepared as described in Example 1 was immobilized on 200 ul Seradyn magnetic beads. The lower strand was removed by incubation of the DNA with 200 ul of 0.1 M NaOH (pH=13) for 3 minutes, the supernatant was removed and an additional washing of 30 seconds with 200 ul of 0.1 M NaOH was performed. Supernatant was removed and the beads were neutralized with 200 ul of 10 mM Tris (pH=7.5), 100 mM NaCl. 2 additional washes with 200 ul NEB2 buffer 2, containing 10 mM Tris (pH=7.9), 50 mM NaCl, 10 mM MgCl2 and 1 mM dithiothreitol, were

performed. After immobilization, the amount of ssDNA was estimated on a 5% PAGE-UREA gel.

About 0.8 ug ssDNA was recovered and incubated in 100 ul NEB2 buffer 2 containing 80 molar fold excess of an equimolar mix of ON_Lam1aB7, ON_Lam2aB7, ON_Lam3lB7 and ON_Lam3rB7 [each oligo in 20 fold molar excess] (see Table 31).

The mixture was incubated at 95° C for 5 minutes and then slowly cooled down to 50° C over a period of 30 minutes. Excess of oligonucleotide was washed away with 2 washes of 200 ul of NEB buffer 2. 4 U/ug of *Hinf I* was added and incubated for 1 hour at 50° C. Beads were mixed every 10 minutes.

After incubation the sample was purified over a Qiagen PCR purification column and was subsequently analysed on a 5% PAGE-urea gel (see FIG. 13A, cleavage was more than 70% efficient).

A schematic of the ligation of the cleaved light chains is shown in FIG. 12B. A mix of bridge/extender pairs was prepared from the Brg/Ext oligo's listed in Table 31 (total molar excess 100 fold) in 1000 U of T4 DNA Ligase (NEB) and incubated overnight at 16° C. After ligation of the DNA, the excess oligonucleotide was removed with a Qiagen PCR purification column and ligation was checked on a Urea-PAGE gel (see FIG. 13B; ligation was more than 95% efficient).

Multiple PCRs were performed containing 10 ng of the ligated material in an 50 ul PCR reaction using 25 pMol ON lamPlePCR and 25 pmol of an equimolar mix of Hu-CL2AscI/HuCL7AscI primer (see Example 1).

PCR was performed at 60° C for 15 cycles using Pfu polymerase. About 1 ug of dsDNA was recovered

per PCR (see FIG. 13C) and cleaved with *ApaL1* and *AscI* for cloning the lambda light chains in pCES2.

Example 5: Capture of human heavy-chain CDR3 population.

A schematic of the cleavage and ligation of antibody light chains is shown in FIGs. 14A and 14B.

Approximately 3 ug of human heavy-chain (IgM) gene RACE material with biotin attached to 5'-end of lower strand was immobilized on 300 uL of Seradyn magnetic beads. The upper strand was removed by washing the DNA with 2 aliquots 300 uL of 0.1 M NaOH (pH 13) for 3 minutes for the first aliquot followed by 30 seconds for the second aliquot. The beads were neutralized with 300 uL of 10 mM Tris (pH 7.5) 100 mM NaCl. The REdaptors (oligonucleotides used to make single-stranded DNA locally double-stranded) shown in Table 32 were added in 30 fold molar excess in 200 uL of NEB buffer 4 (50 mM Potasium Acetate, 20 mM Tris-Acetate, 10 mM Magnesuim Acetate, 1 mM dithiothreitol pH 7.9) to the dry beads. The REadaptors were incubated with the single-stranded DNA at 80 °C for 5 minutes then cooled down to 55 °C over 30 minutes. Excess REdaptors were washed away with 2 washes of NEB buffer 4. Fifteen units of HpyCH4III (NEB) were added in NEB buffer 4 and incubated for 1 hour at 55 °C. The cleaved downstream DNA remaining on the beads was removed from the beads using a Qiagen Nucleotide removal column (see FIG. 15).

The Bridge/Extender pairs shown in Table 33 were added in 25 molar excess along with 1200 units of T4 DNA ligase and incubated overnight at 16 °C. Excess

Bridge/Extender was removed with a Qiagen PCR purification column. The ligated material was amplified by PCR using primers H43.XAExtPCR2 and Hucumnest shown in Table 34 for 10 cycles with the program shown in Table 35.

The soluble PCR product was run on a gel and showed a band of approximately 500 n, as expected (see FIG. 15B). The DNA was cleaved with enzymes *SfiI* and *NotI*, gel purified, and ligated to similarly cleaved vector PCES1.

Example 6: Description of Phage Display Vector CJRA05, a member of the library built in vector DY3F7.

Table 36 contains an annotated DNA sequence of a member of the library, CJRA05, see FIG. 16. Table 36 is to be read as follows: on each line everything that follows an exclamation mark "!" is a comment. All occurrences of A, C, G, and T before "!" are the DNA sequence. Case is used only to show that certain bases constitute special features, such as restriction sites, ribosome binding sites, and the like, which are labeled below the DNA. CJRA05 is a derivative of phage DY3F7, obtained by cloning an *ApaLI* to *NotI* fragment into these sites in DY3F31. DY3F31 is like DY3F7 except that the light chain and heavy chain genes have been replaced by "stuffer" DNA that does not code for any antibody. DY3F7 contains an antibody that binds streptavidin, but did not come from the present library.

The phage genes start with gene ii and continue with genes x, v, vii, ix, viii, iii, vi, i, and iv. Gene iii has been slightly modified in that

eight codons have been inserted between the signal sequence and the mature protein and the final amino acids of the signal sequence have been altered. This allows restriction enzyme recognition sites *EagI* and *XbaI* to be present. Following gene iv is the phage origin of replication (ori). After ori is bla which confers resistance to ampicillin (ApR). The phage genes and bla are transcribed in the same sense.

After bla, is the Fab cassette (illustrated in FIG. 17) comprising:

a) PlacZ promoter,
b) A first Ribosome Binding Site (RBS1),
c) The signal sequence form M13 iii,
d) An *ApaLI* RERS,
e) A light chain (a kappa L20::JK1 shortened by one codon at the V-J boundary in this case),
f) An *AscI* RERS,
g) A second Ribosome Binding Site (RBS2),
h) A signal sequence, preferably PelB, which contains,
i) An *SfiI* RERS,
j) A synthetic 3-23 V region with diversity in CDR1 and CDR2,
k) A captured CDR3,
l) A partially synthetic J region (FR4 after BstEII),
m) CH1,
n) A NotI RERS,
o) A His6 tag,
p) A cMyc tag,
q) An amber codon,
r) An anchor DNA that encodes the same amino-acid sequence as codons 273 to 424 of M13 iii (as shown in Table 37).

s) Two stop codons,

t) An *AvrII* RERS, and

u) A trp terminator.

The anchor (item r) encodes the same amino-acid sequence as do codons 273 to 424 of M13 iii but the DNA is approximately as different as possible from the wild-type DNA sequence. In Table 36, the III' stump runs from base 8997 to base 9455. Below the DNA, as comments, are the differences with wild-type iii for the comparable codons with "!W.T" at the ends of these lines. Note that Met and Trp have only a single codon and must be left as is. These AA types are rare. Ser codons can be changed at all three base, while Leu and Arg codons can be changed at two.

In most cases, one base change can be introduced per codon. This has three advantages: 1) recombination with the wild-type gene carried elsewhere on the phage is less likely, 2) new restriction sites can be introduced, facilitating construction; and 3) sequencing primers that bind in only one of the two regions can be designed.

The fragment of M13 III shown in CJRA05 is the preferred length for the anchor segment. Alternative longer or shorter anchor segments defined by reference to whole mature III protein may also be utilized.

The sequence of M13 III consists of the following elements: Signal Sequence::Domain 1 (D1)::Linker 1 (L1)::Domain 2 (D2)::Linker 2 (L2)::Domain 3 (D3)::Transmembrane Segment (TM):: Intracellular anchor (IC) (see Table 38).

The pIII anchor (also known as trpIII) preferably consists of D2::L2::D3::TM::IC. Another embodiment for the pIII anchor consists of

D2'::L2::D3::TM::IC (where D2' comprises the last 21 residues of D2 with the first 109 residues deleted). A further embodiment of the pIII anchor consists of D2'(C>S)::L2::D3::TM::IC (where D2'(C>S) is D2' with the single C converted to S), and d) D3::TM::IC.

Table 38 shows a gene fragment comprising the *NotI* site, His6 tag, cMyc tag, an amber codon, a recombinant enterokinase cleavage site, and the whole of mature M13 III protein. The DNA used to encode this sequence is intentionally very different from the DNA of wild-type gene iii as shown by the lines denoted "W.T." containing the w.t. bases where these differ from this gene. III is divided into domains denoted "domain 1", "linker 1", "domain 2", "linker 2", "domain 3", "transmembrane segment", and "intracellular anchor".

Alternative preferred anchor segments (defined by reference to the sequence of Table 38) include:

codons 1-29 joined to codons 104-435, deleting domain 1 and retaining linker 1 to the end;

codons 1-38 joined to codons 104-435, deleting domain 1and retaining the rEK cleavage site plus linker 1 to the end from III;

codons 1-29 joined to codons 236-435, deleting domain 1, linker 1, and most of domain 2 and retaining linker 2 to the end;

codons 1-38 joined to codons 236-435, deleting domain 1, linker 1, and most of domain 2 and retaining linker 2 to the end and the rEK cleavage site;

codons 1-29 joined to codons 236-435 and changing codon 240 to Ser(*e.g.*, agc), deleting domain 1, linker 1, and most of domain 2 and retaining linker 2 to the end; and

codons 1-38 joined to codons 236-435 and changing codon 240 to Ser(*e.g.*, agc), deleting domain 1, linker 1, and most of domain 2 and retaining linker 2 to the end and the rEK cleavage site.

The constructs would most readily be made by methods similar to those of Wang and Wilkinson (Biotechniques 2001: 31(4)722-724) in which PCR is used to copy the vector except the part to be deleted and matching restriction sites are introduced or retained at either end of the part to be kept. Table 39 shows the oligonucleotides to be used in deleting parts of the III anchor segment. The DNA shown in Table 38 has an *NheI* site before the DINDDRMA recombinant enterokinase cleavage site (rEKCS). If *NheI* is used in the deletion process with this DNA, the rEKCS site would be lost. This site could be quite useful in cleaving Fabs from the phage and might facilitate capture of very high-afffinity antibodies. One could mutagenize this sequence so that the *NheI* site would follow the rEKCS site, an Ala Ser amino-acid sequence is already present. Alternatively, one could use *SphI* for the deletions. This would involve a slight change in amino acid sequence but would be of no consequence.

Example 7 : Selection of antigen binders from an enriched library of human antibodies using phage vector DY3F31.

In this example the human antibody library used is described in de Haard et al., (Journal of Biological Chemistry, 274 (26): 18218-30 (1999). This library, consisting of a large non-immune human Fab phagemid library, was first enriched on antigen, either

on streptavidin or on phenyl-oxazolone (phOx). The methods for this are well known in the art. Two preselected Fab libraries, the first one selected once on immobilized phOx-BSA (R1-ox) and the second one selected twice on streptavidin (R2-strep), were chosen for recloning.

These enriched repertoires of phage antibodies, in which only a very low percentage have binding activity to the antigen used in selection, were confirmed by screening clones in an ELISA for antigen binding. The selected Fab genes were transferred from the phagemid vector of this library to the DY3F31 vector via *ApaL1-Not1* restriction sites.

DNA from the DY3F31 phage vector was pretreated with ATP dependent DNAse to remove chromosomal DNA and then digested with *ApaL1* and *Not1*. An extra digestion with *AscI* was performed in between to prevent self-ligation of the vector. The *ApaL1/NotI* Fab fragment from the preselected libraries was subsequently ligated to the vector DNA and transformed into competent XL1-blue MRF' cells.

Libraries were made using vector:insert ratios of 1:2 for phOx-library and 1:3 for STREP library, and using 100 ng ligated DNA per 50 µl of electroporation-competent cells (electroporation conditions : one shock of 1700 V, 1 hour recovery of cells in rich SOC medium, plating on amplicillin-containing agar plates).

This transformation resulted in a library size of 1.6×10^6 for R1-ox in DY3F31 and 2.1×10^6 for R2-strep in DY3F31. Sixteen colonies from each library were screened for insert, and all showed the correct size insert (±1400 bp) (for both libraries).

Phage was prepared from these Fab libraries as follows. A representative sample of the library was inoculated in medium with ampicillin and glucose, and at OD 0.5, the medium exchanged for ampicillin and 1 mM IPTG. After overnight growth at 37 °C, phage was harvested from the supernatant by PEG-NaCl precipitation. Phage was used for selection on antigen. R1-ox was selected on phOx-BSA coated by passive adsorption onto immunotubes and R2-strep on streptavidin coated paramagnetic beads (Dynal, Norway), in procedures described in de Haard et. al. and Marks et. al., Journal of Molecular Biology, 222(3): 581-97 (1991). Phage titers and enrichments are given in Table 40.

Clones from these selected libraries, dubbed R2-ox and R3-strep respectively, were screened for binding to their antigens in ELISA. 44 clones from each selection were picked randomly and screened as phage or soluble Fab for binding in ELISA. For the libraries in DY3F31, clones were first grown in 2TY-2% glucose-50 µg/ml AMP to an OD600 of approximately 0.5, and then grown overnight in 2TY-50 µg/ml AMP +/- 1mM IPTG. Induction with IPTG may result in the production of both phage-Fab and soluble Fab. Therefore the (same) clones were also grown without IPTG. Table 41 shows the results of an ELISA screening of the resulting supernatant, either for the detection of phage particles with antigen binding (Anti-M13 HRP = anti-phage antibody), or for the detection of human Fabs, be it on phage or as soluble fragments, either with using the anti-myc antibody 9E10 which detects the myc-tag that every Fab carries at the C-terminal end of the heavy chain followed by a HRP-labeled rabbit-anti-Mouse serum (column 9E10/RAM-HRP), or with

anti-light chain reagent followed by a HRP-labeled goat-anti-rabbit antiserum(anti-CK/CL Gar-HRP).

The results shows that in both cases antigen-binders are identified in the library, with as Fabs on phage or with the anti-Fab reagents (Table 41). IPTG induction yields an increase in the number of positives. Also it can be seen that for the phOx-clones, the phage ELISA yields more positives than the soluble Fab ELISA, most likely due to the avid binding of phage. Twenty four of the ELISA-positive clones were screened using PCR of the Fab-insert from the vector, followed by digestion with *BstNI*. This yielded 17 different patterns for the phOx-binding Fab's in 23 samples that were correctly analyzed, and 6 out of 24 for the streptavidin binding clones. Thus, the data from the selection and screening from this pre-enriched non-immune Fab library show that the DY3F31 vector is suitable for display and selection of Fab fragments, and provides both soluble Fab and Fab on phage for screening experiments after selection.

Example 8: Selection of Phage-antibody libraries on streptavidin magnetic beads.

The following example describes a selection in which one first depletes a sample of the library of binders to streptavidin and optionally of binders to a non-target (*i.e.*, a molecule other than the target that one does not want the selected Fab to bind). It is hypothesized that one has a molecule, termed a "competitive ligand", which binds the target and that an antibody which binds at the same site would be especially useful.

For this procedure Streptavidin Magnetic Beads (Dynal) were blocked once with blocking solution (2% Marvel Milk, PBS (pH 7.4), 0.01% Tween-20 ("2%MPBST")) for 60 minutes at room temperature and then washed five times with 2%MPBST. 450 µL of beads were blocked for each depletion and subsequent selection set.

Per selection, 6.25 µL of biotinylated depletion target (1 mg/mL stock in PBST) was added to 0.250 mL of washed, blocked beads (from step 1). The target was allowed to bind overnight, with tumbling, at 4°C. The next day, the beads are washed 5 times with PBST.

Per selection, 0.010 mL of biotinylated target antigen (1 mg/mL stock in PBST) was added to 0.100 mL of blocked and washed beads (from step 1). The antigen was allowed to bind overnight, with tumbling, at 4°C. The next day, the beads were washed 5 times with PBST.

In round 1, 2 X 10^{12} up to 10^{13} plaque forming units (pfu) per selection were blocked against non-specific binding by adding to 0.500 mL of 2%MPBS (=2%MPBST without Tween) for 1 hr at RT (tumble). In later rounds, 1011 pfu per selection were blocked as done in round 1.

Each phage pool was incubated with 50 µL of depletion target beads (final wash supernatant removed just before use) on a Labquake rotator for 10 min at room temperature. After incubation, the phage supernatant was removed and incubated with another 50 µL of depletion target beads. This was repeated 3 more times using depletion target beads and twice using blocked streptavidin beads for a total of 7 rounds of

depletion, so each phage pool required 350 µL of depletion beads.

A small sample of each depleted library pool was taken for titering. Each library pool was added to 0.100 mL of target beads (final wash supernatant was removed just before use) and allowed to incubate for 2 hours at room temperature (tumble).

Beads were then washed as rapidly as possible (*e.g.*,3 minutes total) with 5 X 0.500 mL PBST and then 2X with PBS. Phage still bound to beads after the washing were eluted once with 0.250 mL of competitive ligand (~1 µµM) in PBST for 1 hour at room temperature on a Labquake rotator. The eluate was removed, mixed with 0.500 mL Minimal A salts solution and saved. For a second selection, 0.500 mL 100 mM TEA was used for elution for 10 min at RT, then neutralized in a mix of 0.250 mL of 1 M Tris, pH 7.4 + 0.500 mL Min A salts.

After the first selection elution, the beads can be eluted again with 0.300 mL of non-biotinylated target (1 mg/mL) for 1 hr at RT on a Labquake rotator. Eluted phage are added to 0.450 mL Minimal A salts.

Three eluates (competitor from 1st selection, target from 1st selection and neutralized TEA elution from 2nd selection) were kept separate and a small aliquot taken from each for titering. 0.500 mL Minimal A salts were added to the remaining bead aliquots after competitor and target elution and after TEA elution. Take a small aliquot from each was taken for tittering.

Each elution and each set of eluted beads was mixed with 2X YT and an aliquot (*e.g.*, 1 mL with 1. E 10/mL) of XL1-Blue MRF' E. coli cells (or other F' cell line) which had been chilled on ice after having been grown to mid-logarithmic phase, starved and

concentrated (see procedure below - "Mid-Log prep of XL-1 blue MRF' cells for infection").

After approximately 30 minutes at room temperature, the phage/cell mixtures were spread onto Bio-Assay Dishes (243 X 243 X 18 mm, Nalge Nunc) containing 2XYT, 1mM IPTG agar. The plates were incubated overnight at 30°C. The next day, each amplified phage culture was harvested from its respective plate. The plate was flooded with 35 mL TBS or LB, and cells were scraped from the plate. The resuspended cells were transferred to a centrifuge bottle. An additional 20 mL TBS or LB was used to remove any cells from the plate and pooled with the cells in the centrifuge bottle. The cells were centrifuged out, and phage in the supernatant was recovered by PEG precipitation. Over the next day, the amplified phage preps were titered.

In the first round, two selections yielded five amplified eluates. These amplified eluates were panned for 2-3 more additional rounds of selection using ~1. E 12 input phage/round. For each additional round, the depletion and target beads were prepared the night before the round was initiated.

For the elution steps in subsequent rounds, all elutions up to the elution step from which the amplified elution came from were done, and the previous elutions were treated as washes. For the bead infection amplified phage, for example, the competitive ligand and target elutions were done and then tossed as washes (see below). Then the beads were used to infect E. coli. Two pools, therefore, yielded a total of 5 final elutions at the end of the selection.

1st selection set

A. Ligand amplified elution: elute w/ ligand for 1 hr, keep as elution

B. Target amplified elution: elute w/ ligand for 1 hr, toss as wash elute w/ target for 1 hr, keep as elution

C. Bead infect. amp. elution: elute w/ ligand for 1 hr, toss as wash elute w/ target for 1 hr, toss as wash elute w/ cell infection, keep as elution

2nd selection set

A. TEA amplified elution; elute w/ TEA 10min, keep as elution

B. Bead infect. amp. elution; elute w/ TEA 10min, toss as wash elute w/ cell infection, keep as elution

Mid-log prep of XL1 blue MRF' cells for infection
(based on Barbas et al. Phage Display manual procedure)

Culture XL1 blue MRF' in NZCYM (12.5 mg/mL tet) at 37°C and 250 rpm overnight. Started a 500 mL culture in 2 liter flask by diluting cells 1/50 in NZCYM/tet (10 mL overnight culture added) and incubated at 37°C at 250 rpm until OD600 of 0.45 (1.5-2 hrs) was reached. Shaking was reduced to 100 rpm for 10 min. When OD600 reached between 0.55-0.65, cells were transferred to 2 x 250 mL centrifuge bottles, centrifuged at 600 g for 15 min at 4°C. Supernatant

was poured off. Residual liquid was removed with a pipette.

The pellets were gently resuspended (not pipetting up and down) in the original volume of 1 X Minimal A salts at room temp. The resuspended cells were transferred back into 2-liter flask, shaken at 100 rpm for 45 min at 37°C. This process was performed in order to starve the cells and restore pili. The cells were transferred to 2 x 250 mL centrifuge bottles, and centrifuged as earlier.

The cells were gently resuspended in ice cold Minimal A salts (5 mL per 500 mL original culture). The cells were put on ice for use in infections as soon as possible.

The phage eluates were brought up to 7.5 mL with 2XYT medium and 2.5 mL of cells were added. Beads were brought up to 3 mL with 2XYT and 1 mL of cells were added. Incubated at 37oC for 30 min. The cells were plated on 2XYT, 1 mM IPTG agar large NUNC plates and incubated for 18 hr at 30°C.

Example 9: Incorporation of synthetic region in FR1/3 region.

Described below are examples for incorporating of fixed residues in antibody sequences for light chain kappa and lambda genes, and for heavy chains. The experimental conditions and oligonucleotides used for the examples below have been described in previous examples (*e.g.*, Examples 3 & 4).

The process for incorporating fixed FR1 residues in an antibody lambda sequence consists of 3 steps (see FIG. 18): (1) annealing of single-stranded

DNA material encoding VL genes to a partially complementary oligonucleotide mix (indicated with Ext and Bridge), to anneal in this example to the region encoding residues 5-7 of the FR1 of the lambda genes (indicated with X..X; within the lambda genes the overlap may sometimes not be perfect); (2) ligation of this complex; (3) PCR of the ligated material with the indicated primer ('PCRpr') and for example one primer based within the VL gene. In this process the first few residues of all lambda genes will be encoded by the sequences present in the oligonucleotides (Ext., Bridge or PCRpr). After the PCR, the lambda genes can be cloned using the indicated restriction site for ApaLI.

The process for incorporating fixed FR1 residues in an antibody kappa sequence (FIG. 19) consists of 3 steps : (1) annealing of single-stranded DNA material encoding VK genes to a partially complementary oligonucleotide mix (indicated with Ext and Bri), to anneal in this example to the region encoding residues 8-10 of the FR1 of the kappa genes (indicated with X..X; within the kappa genes the overlap may sometimes not be perfect) ; (2) ligation of this complex; (3) PCR of the ligated material with the indicated primer ('PCRpr') and for example one primer based within the VK gene. In this process the first few (8) residues of all kappa genes will be encode by the sequences present in the oligonucleotides (Ext., Bridge or PCRpr.). After the PCR, the kappa genes can be cloned using the indicated restriction site for ApaLI.

The process of incorporating fixed FR3 residues in a antibody heavy chain sequence (FIG. 20) consists of 3 steps : (1) annealing of single-stranded DNA material encoding part of the VH genes (for example encoding FR3, CDR3 and FR4 regions) to a partially

complementary oligonucleotide mix (indicated with Ext and Bridge), to anneal in this example to the region encoding residues 92-94 (within the FR3 region) of VH genes (indicated with X..X; within the VH genes the overlap may sometimes not be perfect); (2) ligation of this complex; (3) PCR of the ligated material with the indicated primer ('PCRpr') and for example one primer based within the VH gene (such as in the FR4 region). In this process certain residues of all VH genes will be encoded by the sequences present in the oligonucleotides used here, in particular from PCRpr (for residues 70-73), or from Ext/Bridge oligonucleotides (residues 74-91). After the PCR, the partial VH genes can be cloned using the indicated restriction site for *XbaI*.

It will be understood that the foregoing is only illustrative of the principles of this invention and that various modifications can be made by those skilled in the art without departing from the scope of and sprit of the invention.